



# Macroevolution of gastric *Helicobacter* species unveils interspecies admixture and time of divergence

Annemieke Smet<sup>1,2</sup> · Koji Yahara<sup>3</sup> · Mirko Rossi<sup>4</sup> · Alfred Tay<sup>5</sup> · Steffen Backert<sup>6</sup> · Ensser Armin<sup>7</sup> · James G. Fox<sup>8</sup> · Bram Flahou<sup>2</sup> · Richard Ducatelle<sup>2</sup> · Freddy Haesebrouck<sup>2</sup> · Jukka Corander<sup>9,10,11</sup>

Received: 8 November 2017 / Revised: 29 January 2018 / Accepted: 20 March 2018 / Published online: 25 June 2018  
© International Society for Microbial Ecology 2018

## Abstract

Since the discovery of the human pathogen *Helicobacter pylori*, various other *Helicobacter* species have been identified in the stomach of domesticated and wild mammals. To better understand the evolutionary history of these ecologically similar but genetically distinct species, we analyzed 108 gastric *Helicobacter* genomes and included 54 enterohepatic *Helicobacter* genomes for comparison purposes. An admixture analysis supported the presence of an ecological barrier, preventing the genetic exchange between the gastric and enterohepatic *Helicobacter* species, and unraveled many gene flow events within and across species residing in the stomach. As pets can be colonized by multiple gastric *Helicobacter* species, the genetic exchange between the canine and feline strains was evident, with *H. heilmannii* and *H. bizzozeronii* showing the highest interspecies recombination. An admixture between *H. pylori* (in particular, the ancestral African strains), *H. acinonychis* from wild felines and *H. ceterum* from marine mammals was also identified. Because these latter species do not share the same host, this phenomenon is most likely a remaining signal of shared ancestry. A reconstruction of the time of divergence of the gastric *Helicobacter* spp. revealed that the domestic animal-related *Helicobacter* species evolved in parallel with *H. pylori* and its two closest relatives (*H. acinonychis* and *H. ceterum*), rather than together.

---

These authors contributed equally: Annemieke Smet, Koji Yahara, Mirko Rossi, Freddy Haesebrouck, Jukka Corander.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41396-018-0199-5>) contains supplementary material, which is available to authorized users.

✉ Annemieke Smet  
annemieke.smet@uantwerpen.be

✉ Koji Yahara  
k-yahara@nih.go.jp

✉ Mirko Rossi  
mirko.rossi@helsinki.fi

<sup>1</sup> Laboratory Experimental Medicine and Pediatrics, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

<sup>2</sup> Department of Pathology, Bacteriology and Avian Diseases, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium

<sup>3</sup> Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Tokyo, Japan

<sup>4</sup> Department of Food Hygiene and Environmental Health, Faculty

## Introduction

The stomach has long been considered as a sterile and hostile environment due to its extreme acidity, but more recently it has been established that it is actually populated by microbes. Some inhabitants, that can infect the stomach for decades, are members of the genus *Helicobacter* [1],

of Veterinary Medicine, University of Helsinki, Helsinki, Finland

<sup>5</sup> The Marshall Centre for Infectious Diseases Research and Training, School of Pathology and Laboratory Medicine, University of Western Australia, Nedlands, Perth, WA, Australia

<sup>6</sup> Department Biology, Division Microbiology, University Erlangen Nuremberg, Erlangen, Germany

<sup>7</sup> Institute of clinical and Molecular Virology, Universitätsklinikum Erlangen, Erlangen, Germany

<sup>8</sup> Division of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>9</sup> Department of Biostatistics, University of Oslo, Oslo, Norway

<sup>10</sup> Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

<sup>11</sup> Wellcome Trust Sanger Institute, Cambridge, UK

which can roughly be divided into gastric and enterohepatic *Helicobacter* species [2]. The gastric group exclusively colonizes the stomach of mammals, whereas the enterohepatic group colonizes the liver or intestinal tracts of animals, ranging from mammals to bird and reptiles [3].

The specificity of gastric *Helicobacter* species to colonize and persist in the stomach is the consequence of a series of adaptations that occurred a long time ago [4].

*H. pylori* is by far the most explored gastric *Helicobacter* species, colonizing the stomach of about half the human population and is predominantly transmitted within families rather than spread epidemically [5–7]. This bacterial species has been linked to a wide variety of gastric pathologies, including adenocarcinoma, which is the fifth most common cancer type worldwide [8]. Comparative genomics approaches for a substantial number of *H. pylori* strains have already revealed numerous insights into the evolution of this pathogen [4, 9–11]. *H. pylori* seems to be approximately as old as anatomically modern humans (at least 100 kya) highlighting that infection by this pathogen post-dated the evolution of humans [4]. This suggests that *H. pylori* has probably been acquired via a single-host jump from an unknown, non-human animal host since the stomachs of many animals are also infected by diverse species of *Helicobacter*. The closest relatives of *H. pylori* are *H. acinonychis* from wild felines, which has arisen through a host jump from humans [12], and *H. cetorum* from dolphins and whales [13]. Genetically much more distinct from *H. pylori* are the *Helicobacter* species naturally colonizing the stomach of domestic animals (including cats, dogs, and pigs) which have been designated as the non-*H. pylori Helicobacter* species (NHPH) group [14]. Particularly, cats and dogs are often infected with multiple species that can be regarded as sympatric (i.e., two bacterial species sharing the same niche and host who frequently encounter one another). Up to 8 different NHPH have been described so far in cats, dogs, and pigs [14, 15, 16], of which several harbor a zoonotic potential. A distinct feature is their association with mucosa-associated lymphoid tissue (MALT) lymphoma, another type of gastric cancer that is less prevalent in humans [14, 16]. Not all NHPH colonize the pets and pigs, for example *H. suncus* colonizes the stomach of house musk shrews [17, 18].

Due to the ongoing discovery of novel species in the stomach of pets [16] and similar disease outcomes caused by different gastric *Helicobacter* species that can infect one or multiple hosts [16], the need to study the evolutionary adaptation of these ecologically similar but genetically distinct species becomes even more important. Available studies involving genomic analyses of animal-related gastric *Helicobacter* species are currently restricted to gene content identity and phylogeny [12, 13, 19–23]. This leaves open questions regarding population structure and gene

flow, intra- and interspecific genetic variability, and the age of speciation. Therefore, to improve our understanding of the evolutionary forces acting on the *Helicobacter* genus, we analyzed 108 gastric *Helicobacter* genomes encompassing multiple strains of available gastric species. In addition, 54 enterohepatic *Helicobacter* genomes from 23 species were also included for comparison. These latter species are ecologically and genetically distinct from the gastric members within the genus *Helicobacter*.

Our study provides new insights into the evolution of gastric *Helicobacter* species, showing comprehensive evidence of admixture between species, as well as estimates of the divergence time when these gastric *Helicobacter* spp. have split from their most recent common ancestor.

## Materials and methods

### Helicobacter strains and genome sequencing

A total of 162 strains, consisting of 108 gastric and 54 enterohepatic *Helicobacter* spp., were included in this study (see Supplementary Table 1). More specifically, strains from 11 gastric and 23 enterohepatic species, as shown in Table 1, were selected. Not all the gastric *Helicobacter* species were included in this study, such as *H. suncus* from house musk shrews, due to the absence of its genome sequence. Another gastric species *H. mustelae*, as well as *H. suncus*, is known to be phylogenetically more related to enterohepatic *Helicobacter* species rather than the gastric species [24, 25], and is indicated as gastrointestinal in Table 1 to reflect the potential to colonize both the stomach and the intestinal tract.

In addition, 10 strains from unknown enterohepatic species were also taken into account. Genome sequences from 77 *Helicobacter* strains were already available from the ftp NCBI database (ftp://ftp.ncbi.nlm.nih.gov/; Supplementary Table 1). The other 85 strains were collected during this study and cultivated in vitro under microaerobic conditions for DNA extraction purposes using the Qiagen (Venlo, The Netherlands) Blood & Tissue kit [16]. Subsequently, the DNA samples were used for whole genome sequencing, as previously described [16]. All genomes, including those from the ftp NCBI database, were subjected to gene finding and automatic annotation using Prokka (rapid prokaryotic annotation; [26]).

### Pangenome, niche-specific gene and phylogenetic reconstruction

The pangenome was estimated using the rapid large-scale prokaryotic pan genome analysis (Roary) tool [27]. Briefly, the annotated proteins from all 162 strains were first used

**Table 1** Overview of the number of strains per *Helicobacter* species (11 gastric, 1 gastrointestinal and 23 enterohepatic species) included in this study

<i>Helicobacter</i> species	Niche	Number of strains
<i>H. pylori</i>	Gastric	33
<i>H. acinonychis</i>	Gastric	7
<i>H. ailurogastricus</i>	Gastric	6
<i>H. baculiformis</i>	Gastric	1
<i>H. bizzozeronii</i>	Gastric	10
<i>H. cetorum complex</i>	Gastric	5
<i>H. cynogastricus</i>	Gastric	1
<i>H. felis</i>	Gastric	22
<i>H. heilmannii</i>	Gastric	7
<i>H. salomonis</i>	Gastric	5
<i>H. suis</i>	Gastric	10
<i>H. mustelae</i>	Gastrointestinal	1
<i>H. cinaedi</i>	Enterohepatic	5
<i>H. fennelliae</i>	Enterohepatic	1
<i>H. muridarum</i>	Enterohepatic	4
<i>H. canis</i>	Enterohepatic	1
<i>H. hepaticus</i>	Enterohepatic	1
<i>H. pullorum</i>	Enterohepatic	3
<i>H. bilis</i>	Enterohepatic	7
<i>H. cholecystus</i>	Enterohepatic	1
<i>H. trogonum</i>	Enterohepatic	4
<i>H. rodentium</i>	Enterohepatic	1
<i>H. typhlonius</i>	Enterohepatic	1
<i>H. mesocricetorum</i>	Enterohepatic	1
<i>H. pametensis</i>	Enterohepatic	3
<i>H. canicola</i>	Enterohepatic	1
<i>H. canadensis</i>	Enterohepatic	2
<i>H. apodemus</i>	Enterohepatic	1
<i>H. aurati</i>	Enterohepatic	1
<i>H. marmotae</i>	Enterohepatic	1
<i>H. winghamensis</i>	Enterohepatic	1
<i>H. equorum</i>	Enterohepatic	1
<i>H. macacae</i>	Enterohepatic	1
<i>H. magdeburgensis</i>	Enterohepatic	1
<i>H. sanguini</i>	Enterohepatic	1
<i>Helicobacter</i> sp. <sup>a</sup>	Enterohepatic	10

<sup>a</sup>species name not yet defined

for a BLASTP all-versus-all sequence similarity search. From the BLASTP output, groups of orthologous proteins were predicted using the MCL software, with paralogue clustering enabled. Core genes were defined as present in  $\geq 99\%$  of the genomes. The rest of the genes were considered as the accessory genome. An overview is shown in Table 2. A pangenome matrix was then created based on the presence or absence of all genetic loci in each individual genome. Subsequently, niche-associated genes (gastric

**Table 2** The *Helicobacter* genus pangenome

Roary output	Number of genes
Core genes ( $99\% \leq$ strains $\leq 100\%$ )	399
Soft core genes ( $95\% \leq$ strains $< 99\%$ )	89
Shell genes ( $15\% \leq$ strains $< 95\%$ )	1719
Cloud genes ( $0\% \leq$ strains $< 15\%$ )	12,888
Total genes ( $0\% \leq$ strains $\leq 100\%$ )	15,095

versus enterohepatic) were identified by Scoary analysis using the Roary output, based on a 80% BLAST identity cut-off and paralogue clustering enabled [28]. We used HHblits [29, 30] for iterative protein sequence searching of the niche-associated genes.

Finally, a phylogenetic tree was created based on the concatenated core genes. The phylogenetic tree was built using the randomized accelerated maximum likelihood (RAxML) program [31] by applying the -f a, -p 12345, -x 12345, -# 100, -m GTRGAMMA parameters, and visualized using the interactive tree of life (iTOL) software [32].

### Population structure and admixture analysis

To elucidate the population structure of the 162 *Helicobacter* strains, we used the Bayesian analysis of population structure (BAPS) software to identify groups that are genetically divergent. BAPS is a popular tool for studying population structure and admixture (genetic flux between populations). We used BAPS in a hierarchical manner (hierBAPS) to resolve the population structure at a finer scale, after excluding the non-informative singleton SNPs, to find substructures inside the main clusters [33]. The clustering was performed using two hierarchical levels. Deeper clustering did not change the number of clusters, when compared to two-level clustering.

The admixture analysis was conducted using two different approaches:

(1) The admixture analysis was first conducted based on the 47 BAPS clusters (at hierarchical level 2), with the following specifications: 100 iterations to estimate the admixture coefficient for the individuals, 200 reference individuals from each population, and 20 iterations for estimating the admixture coefficient for reference individuals [33–36].

(2) Subsequently, the chromosome painting algorithm was applied to the genome-wide haplotype data using the linkage model (<http://www.paintmychromosomes.com>). This was used to calculate the expected number of chunks of DNA imported from a donor into a recipient genome [11]. The obtained values were summarized into a matrix. Such co-ancestry matrices were created to investigate the blockwise homology between the core genome of the

following subsets: (1) the canine/feline NHPH, (2) *H. pylori* and its two closest relatives (*H. acinonychis* and *H. ceterum*), and (3) *H. heilmannii*, *H. ailurogastricus* and *H. suis*. For the subset with the canine and feline NHPH, we excluded the *H. baculiformis* and *H. cynogastricus* singleton strains because chromosome painting is known to mistakenly infer distinct singleton strains as hybrids. Subsequently, we used these matrices as inputs in the fineSTRUCTURE algorithm to perform model-based clustering using a Bayesian Markov chain Monte Carlo (MCMC) approach. The fineSTRUCTURE algorithm was run for 100,000 iterations (the first 100,000 iterations were discarded as MCMC burn-in), and the thinning interval was specified at 100.

### Date estimates

The least-square dating (LSD) algorithm [37] was applied to estimate the coalescence of different core genome subsets (generated with Roary), more specifically that of the Hp (including *H. pylori*, *H. acinonychis*, and *H. ceterum*) and canine/feline NHPH groups. A root-to-tip analysis, implemented in TempEst [38], was first applied on each subset for investigating the temporal signal and clock-likeness of molecular phylogenies. Subsequently, date estimates of time to the most recent common ancestor (TMRCA) for each subset were calculated using the LSD software with the following parameters: a mutation rate  $r$  ranging approximately between  $1.8 \times 10^{-7}$  and  $3.6 \times 10^{-7}$  (mean:  $2.6 \times 10^{-7}$ ; s.d.:  $8 \times 10^{-7}$ ; 95%,  $1.28 \times 10^{-7}$ – $3.91 \times 10^{-7}$ ) based on the long-term population-based mutation rate in the human pathogen *H. pylori* [4]; -d datefile (see Supplementary Table 1), -c, -v 2, -r a, -f 100 and -s (sequence length of the alignment file). A 95% confidence interval was also generated.

## Results

### Pangenome reconstruction, niche-specific accessory genome distribution and core genome-based phylogeny

In total, 15,095 orthologous gene clusters were identified of which 399, designated as the core genome, were present among all *Helicobacter* strains (Table 1). The investigation of accessory genomic signals for niche-specialization (i.e., gastric versus enterohepatic specialization) revealed that the larger part of the *Helicobacter* accessory gene pool is specific to enterohepatic species (4416, ~30%) and only a minor fraction (1365, ~9.2%) is specific to *Helicobacter* species with gastric tropism (Supplementary Figure 1a). Of this accessory gene pool, 44 and 7 genes are core in gastric and enterohepatic *Helicobacter* species, respectively. These

two sets of genes unveiled differences in biological processes between both *Helicobacter* groups that could be related to their niche specialization. The genes are listed in Tables 3 & 4. The majority of the gastric core genes (38 out of 44) were absent in *H. mustelae* (Table 3), whereas half of the enterohepatic core genes (3 out of 7) were present (Table 4).

The division between enterohepatic and gastric *Helicobacter* species is also evident in the phylogenetic reconstruction based on the *Helicobacter* core genome, which resulted in two separate monophyletic groups (Fig. 1). The gastric helicobacters, characterized by long internal and short external branches, could be further divided into the Hp (including *H. pylori* and its two closest relatives, *H. acinonychis* and *H. ceterum*) and NHPH (including the canine, feline, and porcine helicobacters) clades, in which each species was clearly represented by a monophyletic group (Fig. 1). On the contrary, the enterohepatic clade appears to be rather star-like (i.e., short internal and long external branches), suggesting either a rapid ancient radiation event or pervasive recombination among different species. *H. mustelae*, which has been associated with gastritis, peptic ulcers, MALT lymphoma, and adenocarcinoma in domestic ferrets [20, 24, 25], clustered within the clade of the enterohepatic *Helicobacter* species based on the core genomic variation (Fig. 1). This phenomenon was also reflected in the accessory genome, where it can be seen that *H. mustelae* shares more genes with the enterohepatic *Helicobacter* species than with the gastric species (Tables 3 and 4, Supplementary Figure 1b). Although enteric colonization has not yet been described, *H. mustelae* has been detected in high amounts in feces of young ferrets, suggesting lower bowel colonization or transit of the organism from its gastric niche [24, 25]. The above findings thus confirm previous hypotheses [16, 18], which emphasize the capability of *H. mustelae* to colonize both the stomach and the intestinal tract.

### Population structure of the genus *Helicobacter*

We estimated the number of populations by grouping the 162 strains into genetically divergent clusters using BAPS [33]. Overall, the population assignment of the *Helicobacter* genomes was well correlated with the different clades of the core genome-based phylogenetic tree (Fig. 1). In total 16 clusters were identified at the first hierarchical level (first colored bar in Fig. 1 & Supplementary Table S2), where each gastric species was designated as a different population, with the exception of *H. acinonychis* (from wild felines), *H. cynogastricus* (singleton from a dog), and *H. baculiformis* (singleton from a cat). These latter species clustered with *H. pylori* (in particular, the ancestral African strains), *H. felis* (from dogs and cats), and *H. salomonis*

**Table 3** Functional annotation of core genes of gastric *Helicobacter* species

Gene name	Annotation <sup>c</sup>	Functional categories
<i>hbpA</i>	Heme-binding protein A precursor	Heme metabolism
<i>dppB</i>	Dipeptide transport system permease protein DppB	Peptide transport
<i>dppC</i>	Dipeptide transport system permease protein DppC	Peptide transport
<i>oppF</i>	Oligopeptide transport ATP-binding protein OppF	Peptide transport
<i>oppD</i> <sup>a,b</sup>	Oligopeptide transport ATP-binding protein OppD	Peptide transport
<i>nixA</i>	High-affinity nickel-transport protein NixA	Nickel transport
<i>hpn</i>	Nickel-binding protein with high content in histidine Hpn	Nickel transport
<i>trpGD</i> <sup>a,b</sup>	Bifunctional protein TrpGD	Tryptophan metabolism
<i>trpC</i> <sup>a</sup>	Tryptophan biosynthesis protein TrpCF	Tryptophan metabolism
<i>hpaA</i> <sup>a,b</sup>	Neuraminylactose-binding hemagglutinin precursor (NLBH)	Binding to sialic acid
<i>unknown</i> <sup>a</sup>	ribonuclease BN/unknown domain fusion protein	Endonuclease activity
<i>ptlH</i> <sup>a</sup>	Type IV secretion system protein PtlH	Type IV secretory system
<i>clsA</i> <sup>a</sup>	Major cardiolipin synthase ClsA	Lipid metabolism
<i>dadA</i> <sup>a</sup>	D-amino acid dehydrogenase small subunit	Electron transport
<i>hefG</i> <sup>a</sup>	<i>Helicobacter</i> outer membrane efflux protein (Hef family)	Membrane biogenesis
<i>horE</i> <sup>a</sup>	<i>Helicobacter</i> outer membrane protein (Hor family)	Membrane biogenesis
<i>horL</i> <sup>a</sup>	<i>Helicobacter</i> outer membrane protein (Hor family)	Membrane biogenesis
<i>hofC</i> <sup>a</sup>	<i>Helicobacter</i> outer membrane protein (Hof family)	Membrane biogenesis
<i>hofE</i> <sup>*</sup>	<i>Helicobacter</i> outer membrane protein (Hof family)	Membrane biogenesis
<i>hofF</i> <sup>a</sup>	<i>Helicobacter</i> outer membrane protein (Hof family)	Membrane biogenesis
<i>hofH</i> <sup>a</sup>	<i>Helicobacter</i> outer membrane protein	Membrane biogenesis
<i>omp_1</i> <sup>a</sup>	Putative outer membrane protein	Membrane biogenesis
<i>omp_2</i> <sup>a</sup>	Putative outer membrane protein	Membrane biogenesis
<i>omp_3</i> <sup>a</sup>	Putative outer membrane protein	Membrane biogenesis

There are other 20 genes encoding a protein with unknown (hypothetical) function

<sup>a</sup>absent in *H. mustelae*

<sup>b</sup>multiple paralogs present

<sup>c</sup>by Prokka and HHblits

**Table 4** Functional annotation of core genes of enterohepatic *Helicobacter* species

Gene	Annotation <sup>b</sup>	Functional categories
<i>macB_1</i>	Macrolide export ATP-binding/permease protein MacB	ABC transporter system
<i>macB_2</i>	Macrolide export ATP-binding/permease protein MacB	ABC transporter system
<i>yqjA</i>	Inner membrane protein YqjA	Membrane biogenesis
<i>argH1</i> <sup>a</sup>	Argininosuccinate lyase 1	Arginine biosynthesis
<i>argF</i> <sup>a</sup>	Ornithine carbamoyltransferase	Arginine biosynthesis
<i>argG</i> <sup>a</sup>	Argininosuccinate synthase	Arginine biosynthesis

There is another gene encoding a protein with unknown (hypothetical) function

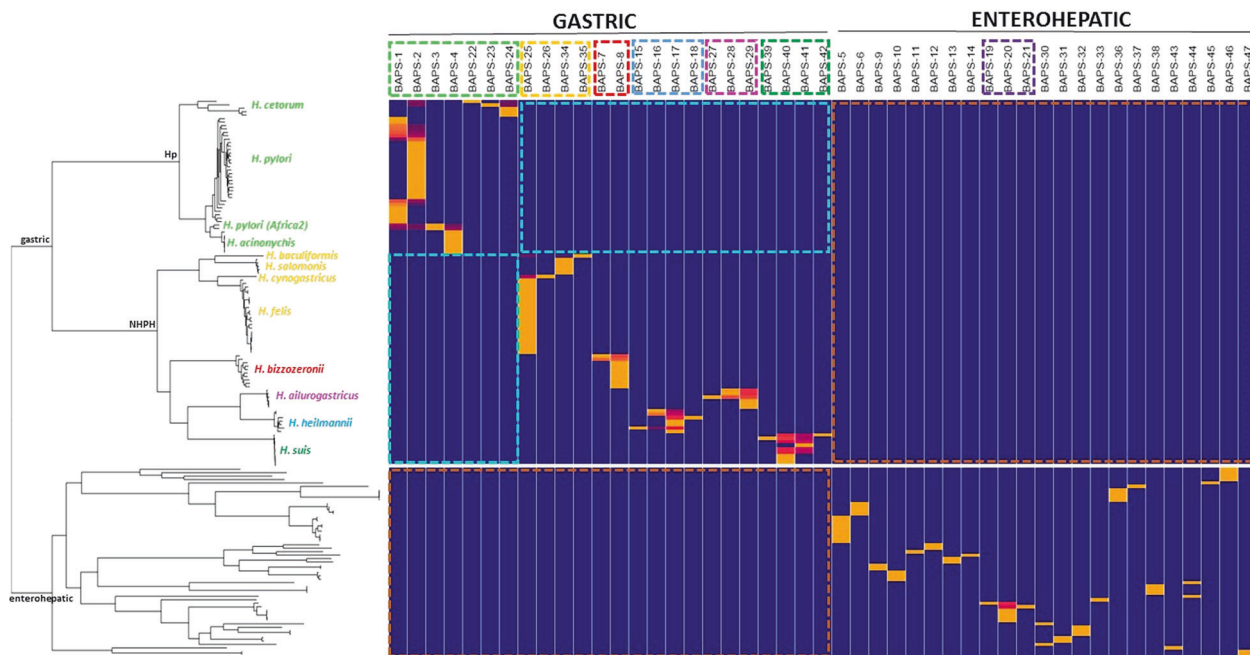
<sup>a</sup>present in *H. mustelae*

<sup>b</sup>by Prokka and HHblits

(from dogs) populations, respectively (Fig. 1; Supplementary Table 2). For the enterohepatic *Helicobacter* spp., however, several populations contained more than one species, as in the case of the population containing *H. mustelae* (Fig. 1; Supplementary Table 2). Clustering at a deeper level (second colored bar in Fig. 1; Supplementary Table 2) further differentiated the *Helicobacter* strains into finer populations (47 in total), as well as reassigned species

into new populations (Fig. 1 & Supplementary Table 2). More specifically, at this finer level of resolution, *H. pylori*, *H. cetorum* (from dolphins and whales), *H. bizzozeronii* (from a cat, dogs, and humans), *H. heilmannii* (from cats), *H. ailurogastricus* (from cats) and *H. suis* (from pigs) were divided into 3, 3, 2, 4, 3, and 3 subpopulations, respectively. Conversely, *H. acinonychis* (from wild felines), *H. cynogastricus* (from a dog), *H. baculiformis* (from a cat),





**Fig. 2** Admixture analysis based on the 47 BAPS predicted population groups at hierarchical level 2 (columns), in association with the *Helicobacter* core genome phylogeny. Each color indicates the estimated percentage of a *Helicobacter* strain to originate from the corresponding BAPS population. The percentages of admixture range from 0% (blue color) to 100% (yellow color), and are also indicated in Supplementary Table 3. The admixture sources for (1) *H. pylori* (BAPS1–3), *H. acinonychis* (BAPS4), and *H. ceterum* (BAPS22 (138563\_11), BAPS23 (138563\_9), and BAPS24) and (2) *H. felis* (BAPS25), *H. cynogastricus* (BAPS26), *H. salomonis* (BAPS34), and *H. baculiformis* (BAPS35) populations are highlighted in green and yellow dashed boxes, respectively. Admixture sources for

*H. bizzoeronii* (BAPS7–8), *H. heilmannii* (BAPS15–18), *H. ailurogastricus* (BAPS27–29), and *H. suis* (BAPS39–42) are highlighted in red, blue, pink, and dark green dashed boxes, respectively. The species names shown in the gastric clade of the core genome tree are presented in the same color, where admixture was identified in the corresponding BAPS clusters. Signatures of admixture between enterohepatic species were only observed between *H. cinaedi* (BAPS19–20) and *H. magdeburgensis* (BAPS21), and are highlighted by a purple dashed box. The absence of admixture between (1) the Hp and NHPH clades and (2) the gastric and enterohepatic groups is highlighted by cyan and brown-red dashed boxes, respectively. The figure is drawn using Phandango; ([59] <https://jameshadfield.github.io/phandango/>)

*H. mustelae* (from a ferret) and several enterohepatic helicobacters were each assigned to a separate population (Fig. 1; Supplementary Table 2). The *H. pylori* and *H. ceterum* subpopulations reflect differences in geographical locations or hosts, whereas the two *H. bizzoeronii* subpopulations could partly correspond to host differences, namely cats versus dogs (Fig. 1). On the contrary, the subpopulations found within *H. heilmannii*, *H. ailurogastricus*, and *H. suis* could not be explained by differences in hosts or geography. The isolates of these latter species were all obtained from a different single animal of the same species (cat or pig) residing in the same country.

### Inference of admixture events supporting interchangeability between species sharing the same niche, and the ecological barrier to genetic exchange between gastric and enterohepatic *Helicobacter* species

As recombination is an important contributor to sequence diversity and population heterogeneity within a bacterial species [39–41], we searched for evidence of intra- and

interspecies genetic exchange events using different approaches.

An admixture analysis was first applied to the 47 hierarchically clustered populations. Figure 2 illustrates a colored map of the *Helicobacter* core genome phylogeny and the 47 BAPS clusters, where each color is equal to the estimated percentage for a *Helicobacter* strain (as shown in Supplementary Table 3) to have ancestry in the corresponding BAPS population. The BAPS admixture analysis revealed patterns of intra-species admixture within *H. pylori*, *H. bizzoeronii*, *H. heilmannii*, *H. ailurogastricus*, and *H. suis* (Fig. 2). Conversely, signatures of interspecies admixture were only observed within the Hp and NHPH clades, and more specifically in the canine *H. cynogastricus* strain (BAPS-26; 34% of its ancestry being derived from the canine/feline *H. felis* (BAPS-25)), the feline *H. baculiformis* strain (BAPS-35, 16 and 6% of its ancestry being derived from the canine/feline *H. felis* (BAPS-25) and the canine *H. salomonis* (BAPS-34), respectively), *H. ceterum* from white-sided dolphins (BAPS-22 & BAPS-23, 21% of its ancestry being derived from the human *H. pylori* Amerindian and East Asian

subpopulations (BAPS-2)) and the human *H. pylori* ancestral Africa 2 population (BAPS-3, 21.5% of its ancestry being derived from the wild feline *H. acinonychis* (BAPS-4)) (Fig. 2; Supplementary Table 3). Interestingly, admixture between members of the Hp and NHPH clades was not observed (dashed cyan rectangle in Fig. 2). In addition, the enterohepatic species were also included in the BAPS model, to explore any potential genetic exchange between species residing in a different niche. However, such a genetic exchange was not found (dashed orange rectangle in Fig. 2), suggesting the existence of barriers to the genetic exchange between these ecologically distinct groups.

Chromosome painting and the fineSTRUCTURE algorithm were then applied to infer the possible genetic exchange among gastric *Helicobacter* species at the strain level. Based on the above BAPS admixture results, the patterns of genetic exchange were further investigated between the canine/feline gastric helicobacters (Fig. 3) and between *H. pylori*, *H. acinonychis*, and *H. cetorum* (Fig. 4). In addition, the porcine *H. suis* species is a sister clade of the feline *H. heilmannii* and *H. ailurogastricus* species (based on the core genome phylogeny; Fig. 1). Therefore, it might be assumed that these species, despite not sharing the same host, are more closely related, when compared to the other members of the NHPH group. The inference of admixture among these three species was therefore also considered (Fig. 5). The co-ancestry matrices, visualized as heatmaps in Figs. 3–5, showed that many events of intra- and interspecies genetic exchange have occurred in both the Hp and NHPH groups.

More specifically, the interspecific genetic exchange was particularly evident among canine and feline gastric *Helicobacter* spp. that can share the same hosts. *H. heilmannii* and *H. bizzozeronii* showed the highest level of interspecies recombination (Fig. 3). Both species have exchanged genetic material with each other, in particular the feline *H. bizzozeronii* strain (56877\_22) with all feline *H. heilmannii* strains (Fig. 3). Moreover, from *H. bizzozeronii*, which also showed the highest intraspecies recombination, DNA was imported into the canine and feline *H. felis*, the feline *H. ailurogastricus*, and the canine *H. salomonis* species, whereas from *H. heilmannii* DNA was additionally imported into the feline *H. ailurogastricus* species (Fig. 3).

Within the Hp clade, *H. cetorum* showed the highest intraspecies admixture (Fig. 4). The analysis also revealed admixture between (1) *H. pylori* (ancestral population Africa 2) and *H. acinonychis*, (2) *H. pylori* (particularly, ancestral Africa 2 and the W-African subpopulation) and *H. cetorum* and (3) *H. acinonychis* and *H. cetorum* (Fig. 4).

Within the NHPH clade, furthermore, a few signatures of genetic exchange from porcine *H. suis* to the feline *H. heilmannii* or *H. ailurogastricus* were noted (Fig. 5).

## Dates of the most recent common ancestor (TMRCA) within the gastric *Helicobacter* clade

The ages of splits between individual lineages within the gastric *Helicobacter* species were inferred by applying fast dating using least-squares criteria and algorithms implemented in the LSD software, which estimates the substitution rate and the dates of all ancestral nodes of a given tree [37]. We first analyzed the date of the split between the most ancient *H. pylori* population (Africa 2) and *H. acinonychis*, and their TMRCA was estimated to be 158,7 kya (95% CI: 157,5–159,7). This is in the range of the previously inferred coalescent dates (within the last 200 kya [12]; 88–116 kya [4]), thus validating our approach. To infer the split between *H. pylori* and *H. cetorum*, we analyzed the core genome alignment of the *H. pylori* Africa 2 population and the *H. cetorum* strains from marine mammals. The split date between *H. cetorum* and *H. pylori* was estimated to be ~610 kya (95% CI: 608,2–612,5) (Fig. 6a).

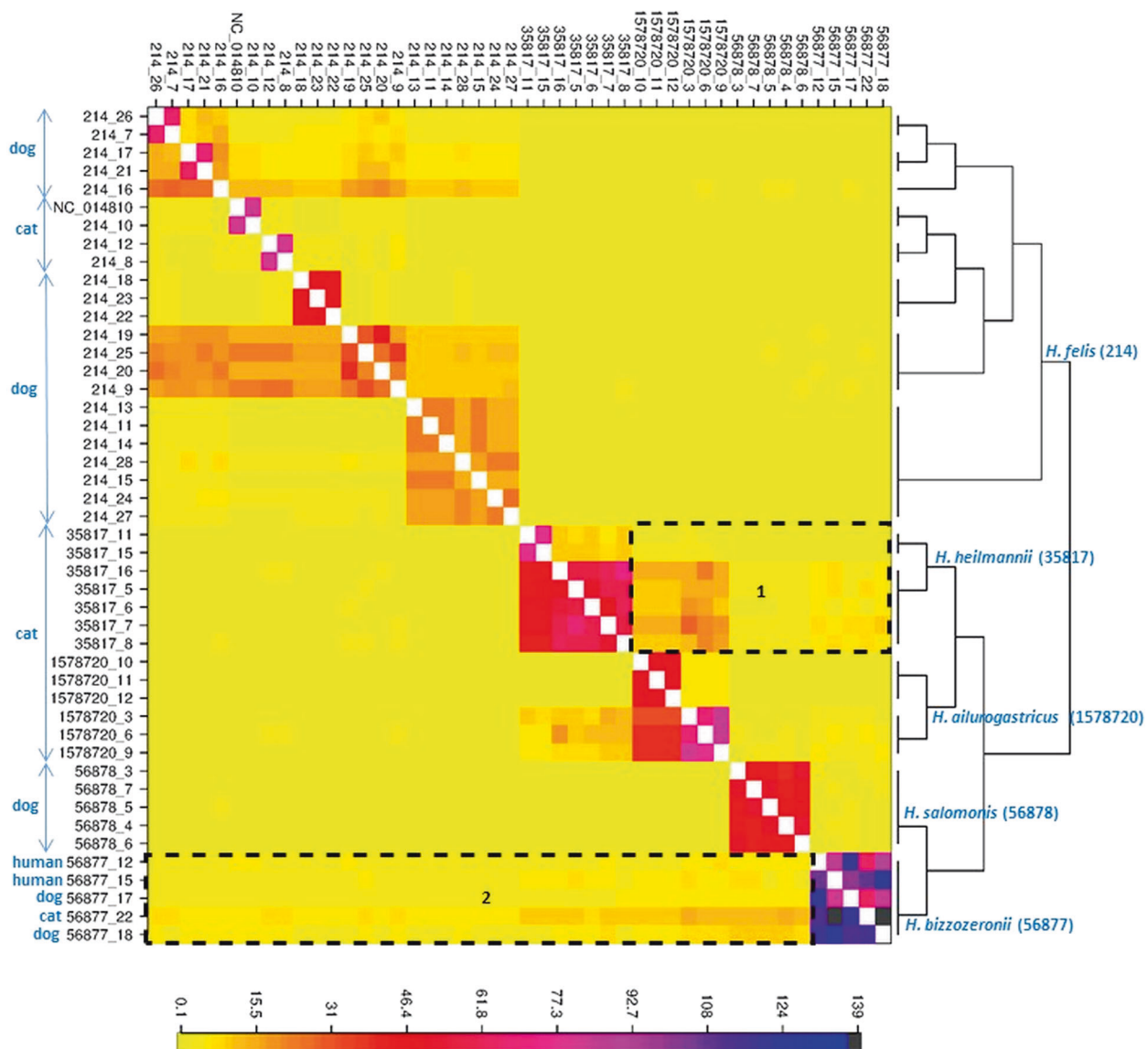
Interestingly, within the NHPH clade, it appeared that speciation occurred much earlier, compared to the date inferred within the *H. pylori* clade. More specifically, the TMRCA for the NHPH species was estimated to be 1,96 Mya [95%\_CI: 1,947–1,967] (Fig. 6b). Moreover, the TMRCA of (1) the *H. felis*, *H. baculiformis*, *H. cynogastricus*, and *H. salomonis* and (2) of the *H. heilmannii* and *H. ailurogastricus* subgroups have been inferred to be 1,52 Mya (95% CI: 1,514–1,525) and 690 kya (95% CI: 687,5–693,5), respectively (Fig. 6b).

## Discussion

The *Helicobacter* genus currently comprises 46 Gram-negative species that have established symbiotic relationships in the gastrointestinal tract of one or more hosts; several of these are of pathogenic importance both to humans and animals [17]. This work aimed to elucidate a scenario of *Helicobacter* evolution and, in particular, the natural history of gastric *Helicobacter* species based on comparative genomics encompassing the species known to date.

Here, we first characterized the *Helicobacter* pangenome at the genus level, to investigate gene variation and phylogeny. The enterohepatic *Helicobacter* species harbor a larger accessory gene pool than the gastric *Helicobacter* species (Supplementary Figure 1), which could be assigned to their larger genome size (enterohepatic: median genome length of ca. 2 Mb versus gastric: median genome length of 1.63 Mb; [16]). However, this could also be due to the fact that the intestinal niche is less hostile to bacteria than the acidic environment in the stomach, and is thus populated with a more luxurious microbiome that allows frequent exchange of genetic features [1, 42]. Identifying





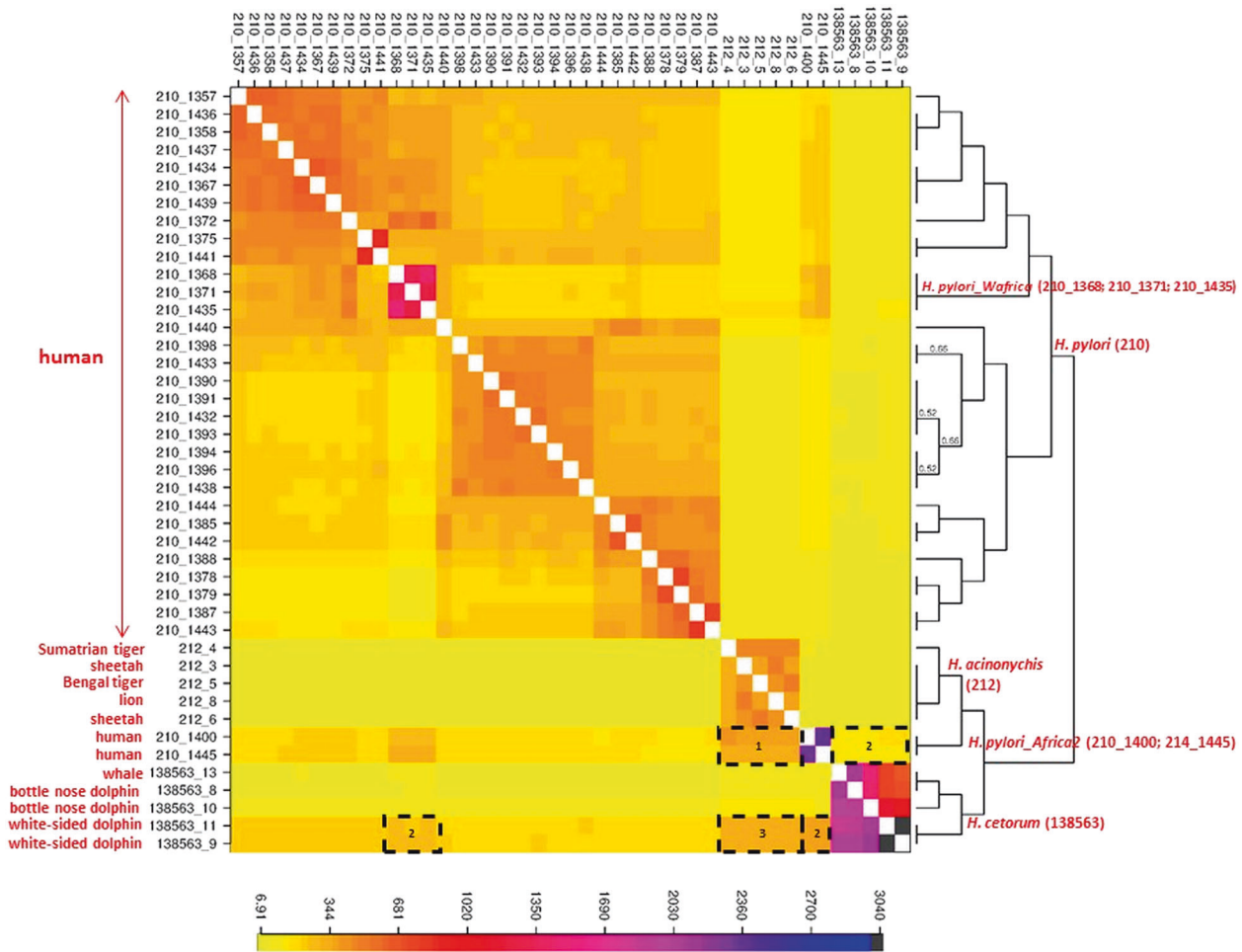
**Fig. 3** Co-ancestry matrix with evidence of genetic flux between *H. felis*, *H. bizzozeronii*, *H. salomonis*, and *H. heilmannii*, based on fineSTRUCTURE analysis. The singleton strains *H. baculiformis* and *H. cynogastricus* were excluded from this analysis. The color of each cell in the matrix indicates the expected number of chunks imported from a donor genome (column) into a recipient genome (row). The

strain IDs (see Supplementary Table 1) and their corresponding host are indicated on the right and the species name is indicated in the tree topology. The highest interspecies recombination seen for *H. heilmannii* (1) and *H. bizzozeronii* (2) is highlighted by dashed black boxes

mechanisms involved in niche adaptation is crucial to understand pathogen evolution [43–45]. In our study, we explored genetic features that distinguish gastric from enterohepatic *Helicobacter* species as in recent studies [46, 47], but by means of more than 3-fold increase of genome sequences of gastric NHPH and enterohepatic species. Genes specific to the gastric *Helicobacter* species were related to nickel homeostasis (e.g., nickel-binding protein Hpn [47]), peptide transport [48], outer membrane biogenesis (particularly outer membrane proteins (OMPs) from the Hor and Hof family; [49]) and tryptophan metabolism. Genes specific to the enterohepatic *Helicobacter* species

were related to resistance to macrolides and the ability to synthesize L-arginine from L-ornithine. Arginine and ornithine are precursors of nitric oxide and polyamines, respectively, and play essential roles in permeability and adaptive responses of the gut [50]. Acquisition of the above traits is likely to be important for the adaptation of *Helicobacter* to the gastric or enterohepatic niche.

Ecological demarcations were also further underlined in the *Helicobacter* core genome phylogeny that resulted in the gastric and enterohepatic clades (Fig. 1). Boundaries between bacterial populations due to ecological segregation could limit recombination events, which play an important



**Fig. 4** Co-ancestry matrix, showing evidence of genetic flux between *H. pylori*, *H. acinonychis*, and *H. cetorum*, based on fineSTRUCTURE analysis. The color of each cell in the matrix indicates the expected number of chunks imported from a donor genome (column) into a recipient genome (row). The strain IDs (see Supplementary Table 1) and their corresponding hosts are indicated on the right and the species

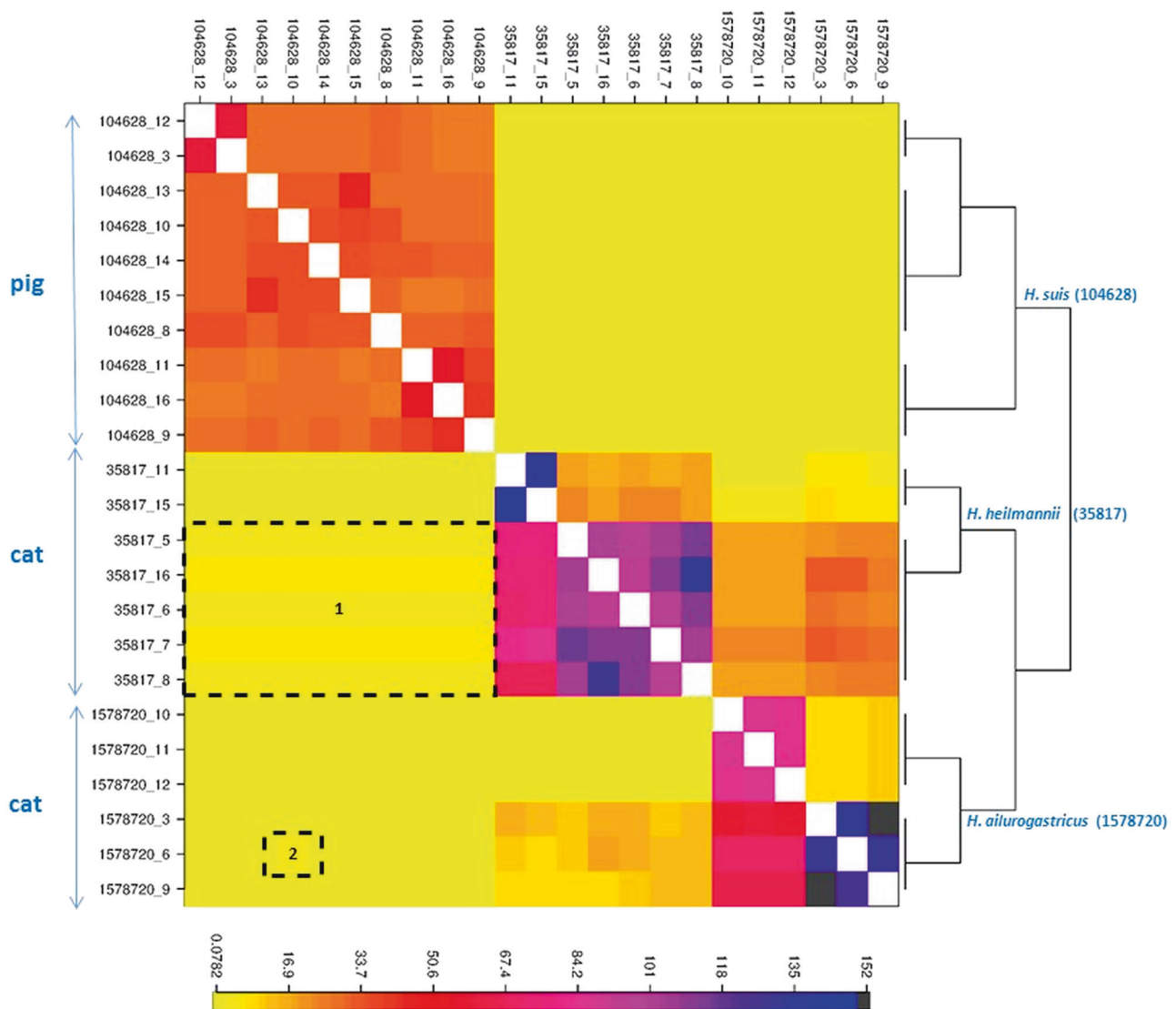
name is indicated in the tree topology. Admixture between (1) *H. pylori* (ancestral population Africa 2) and *H. acinonychis*, (2) *H. pylori* (particularly, ancestral Africa 2 and the W-African subpopulation) and *H. cetorum* and (3) *H. acinonychis* and *H. cetorum*, are highlighted by black dashed boxes

evolutionary role in the speciation process by defining the population structure of bacterial species [40, 41]. Indeed, in our study, the genetic exchange between the gastric and enterohepatic *Helicobacter* species was not observed (Fig. 2), thus supporting the presence of an ecological barrier which could have been taken place when the first part of the digestive tract became specialized (i.e., the stomach with acid production). Nevertheless, a genetic barrier effect could not be ruled out as an alternative explanation. In this latter case, the reduced efficiency of mismatch repair following homologous recombination between divergent sequences may represent a major contributing factor, as shown for *Campylobacter* [34, 44].

For the gastric *Helicobacter* group, however, extensive gene-flow occurred within each gastric species, emphasizing the genetic variability which is a known hallmark of *Helicobacter* [16, 40]. Admixture between species was also

noted. As pets can be colonized by multiple gastric species [14], the genetic exchange among the canine and feline NHPH was evident with *H. heilmannii* and *H. bizzozeronii*, showing the highest level of inter-species recombination (Fig. 3). These two latter species are also the most prevalent species in cats and dogs, respectively [16, 17]. Furthermore, our admixture analysis based on the BAPS population assignment indicated that the singleton pet-associated strains canine *H. cynogastricus* and feline *H. baculiformis* are hybrids, in that they have received a considerable amount of DNA from *H. felis* (Fig. 2).

Signatures of admixture were also observed between closely related species not sharing the same host. These phenomena most likely represent remaining signals of shared ancestry. The signatures of genetic exchange from the porcine *H. suis* to the feline *H. heilmannii* or *H. ailurogastricus* (Fig. 5) might represent the result of shared



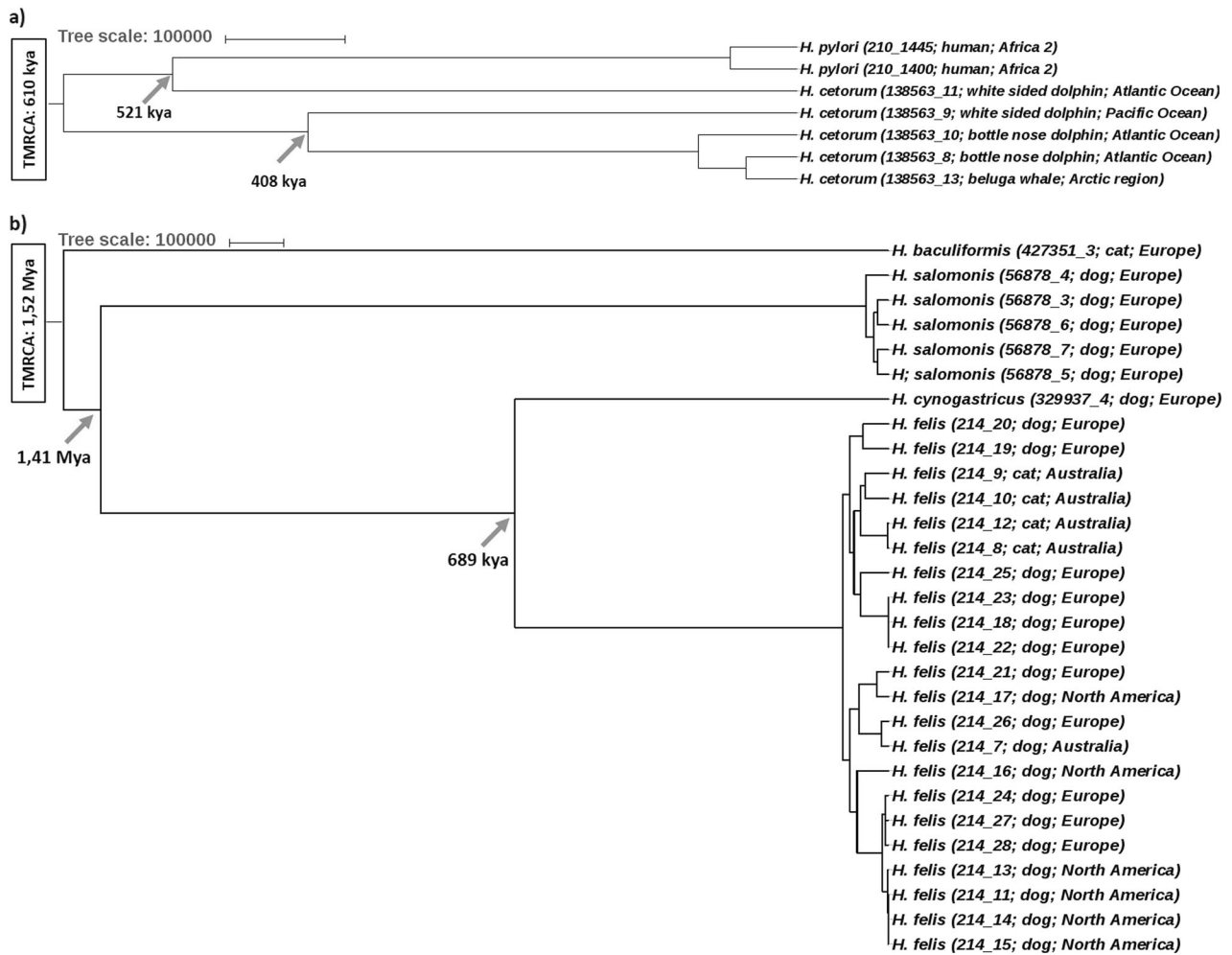
**Fig. 5** Co-ancestry matrix, showing evidence of genetic flux between *H. heilmannii*, *H. ailurogastricus*, and *H. suis*, based on fineSTRUCTURE analysis. The color of each cell in the matrix indicates the expected number of chunks imported from a donor genome (column) into a recipient genome (row). The strain IDs (see

Supplementary Table 1) and their corresponding hosts are indicated on the right and the species name is indicated in the tree topology. DNA imports from *H. suis* into *H. heilmannii* (1) or *H. ailurogastricus* (2) are highlighted by black dashed boxes

genetic metabolic features, since these species have the same in vitro growth requirements, as opposed to the other NHPH [17, 22, 51]. Signals of remaining ancestry were also observed within the Hp clade (Fig. 4). The admixture noted between *H. pylori* ancestral Africa 2 and *H. acinonychis* from wild felines is in agreement with previous findings [4], where it was shown that the progenitors of this ancient *H. pylori* population are the source of *H. acinonychis*. Furthermore, gene flow events were also observed between *H. cetorum* from white-sided dolphins and the human Africa 2 and W-African *H. pylori* based on the fine-structure analysis at the strain level (Fig. 4). The shared ancestry observed between *H. cetorum* and the ancient African *H. pylori* strains might assume that the common

ancestor of both species originates in the African continent as already suggested for *H. pylori* [4].

Elucidating the ancestral ages of splits between individual lineages within the gastric *Helicobacter* phylogeny completes the reconstruction of the historical path of these species [4]. This has already been well examined for *H. pylori* and *H. acinonychis*, where the minimum age of association between *H. pylori* and humans was estimated to be ~100 kya. Furthermore, *H. acinonychis* resulted from a later host jump from humans (ancestral Africa 2) to large felines ca. 43–56 kya [4]. In our study, we found that the coalescent for *H. pylori* (ancestral Africa) and *H. cetorum* dated to ca. 600 kya. In this timeframe, the modern humans and Neanderthals have also diverged [52]. This confirms the



**Fig. 6** Time trees of *Helicobacter* subgroups. **a** The divergence between *H. pylori* (ancestral Africa 2) and *H. cetorum* began 610 kya. **b** Time of divergence among the *H. baculiformis*, *H. salomonis*,

*H. cynogastricus* and *H. felis* subgroups ranged between 1,52 Mya and 689 kya. The scale bar indicates the number of years. Figures are drawn using iTOL

presence of *H. pylori* in humans before domestication, and confirms that the acquisition of *H. pylori* in humans resulted from a jump from an animal host [4]. Furthermore, the evolution of cetaceans had already occurred on the Indian subcontinent 5 Mya [53]. This indicates that *H. cetorum* did not coevolve with marine mammals, but made the jump to white-sided dolphins much earlier in the past (ca. 500 kya), and diverged thereafter to other cetaceans ca. 410 kya. Coevolution between a microbe and its host generally results in decreased pathogenicity, but a disruption caused by jumps between hosts has been associated with an increase in disease severity, as shown for *Staphylococcus aureus* and recently also for *H. suis* [54, 55]. Since infection with *H. cetorum* has been correlated with gastric ulcers in marine mammals [13, 17], these animals are thus probably not the natural host. Whether the origin of *H. cetorum* derives from a marine or terrestrial ecosystem remains to be further elucidated.

The gastric *Helicobacter* species from domestic animals are much more distinct from *H. pylori* and its closest relatives, *H. cetorum* and *H. acinonychis*. Due to the lack of admixture signals between Hp and NHPH, it is likely that these two groups did not evolve together, but rather in parallel, following a very distinct evolutionary path. This hypothesis was confirmed when estimating the time of divergence between the different canine and feline NHPH, indicating an ancient association between these species and their hosts. The most common recent ancestor for different canine/feline NHPH subsets ranged from 1,96 Mya to 690 kya. Taking into account that the common ancestor of cats and dogs is ca. 3–4 Mya [56, 57] and that domestication has occurred more recently [4], our estimations suggest that these pet-associated *Helicobacter* species coevolved with their host far before the domestication of either cats or dogs. Hitherto, these *Helicobacter* species cause little or no harm in cats and dogs [14], highlighting that pets are most

probably their original host. On the contrary, the common ancestor of *H. suis*, the other member of the NHPH group and initially originating from asymptomatic non-human primates, existed until approximately 200 kya (193–197 kya) [55]. This species made a host jump to pigs between 100 and 15 kya where it causes gastric disease and whereby pig domestication has had a significant impact on the spread of *H. suis* in the pig population [55, 58]. As NHPH can occasionally also infect humans, it is thus plausible to assume that their zoonotic potential emerged after domestication of cats, dogs and pigs (ca. 100 kya).

In summary, our data provides new insights into the evolution of gastric *Helicobacter* species, with comprehensive evidence of admixture between species and date estimates for the historical events of gastric *Helicobacter* speciation.

**Acknowledgements** We thank Sofie Debruyckere and Nathalie Van Rysselberghe for their skilled technical assistance. This work was supported by the Research Fund of Ghent University (01G00408), by the Research Foundation Flanders (FWO Vlaanderen, 1220217N), by the Research Fund of Antwerp University (34871) and by “Grant-in-Aid for Scientific Research” from MEXT (15K21554 to KY). The work of SB was supported by a grant of the German Science Foundation (project A04 in CRC-1181). We thank Editage ([www.editage.jp](http://www.editage.jp)) for English language editing. Computational calculations were performed at the Human Genome Center, at the Institute of Medical Science (the University of Tokyo).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Yang I, Nell S, Suerbaum S. Survival in the hostile territory: the microbiota of the stomach. *FEMS Microbiol Rev.* 2013;37:736–61.
- Solnick JV, Schauer DB. Emergence of diverse *Helicobacter* species in the pathogenesis of gastric and enterohepatic diseases. *Clin Microbiol Rev.* 2001;14:59–97.
- Mobley HLT, Mendz GL, Hazell SL. Enterohepatic *Helicobacter* species. *Helicobacter pylori: Physiology and Genetics.* Washington (DC): ASM Press; 2001. 2001Chapter 43
- Moodley Y, Bodo L, Bond RP, Nieuwoudt M, Soodyall H, Schlegelbusch CM, et al. Age of the association between *Helicobacter pylori* and man. *Plos Path.* 2012;8:e1002693.
- Schwarz S, Morelli G, Kusecek B, Manica A, Balloux F, Owen RJ, et al. Horizontal versus familial transmission of *Helicobacter pylori*. *Plos Path.* 2008;4:e1000180.
- Kawai M, Furuta Y, Yahara K, Tsuru T, Oshima K, Handa N, et al. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genome. *BMC Microbiol.* 2011;11:104.
- Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat Rev Microbiol.* 2007;5:441–52.
- Suerbaum S, Michetti P. *Helicobacter pylori* infection. *New Engl J Med.* 2002;347:1175–86.
- Kojima KK, Furuta Y, Yahara K, Fukuyo M, Shiwa Y, Nishiumi S, et al. Population evolution of *Helicobacter pylori* through diversification in DNA methylation and interstrain sequence homogenization. *Mol Biol Evol.* 2016;33:2848–59.
- Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows JL, et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science.* 2016;351:162–5.
- Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, Hattori M, et al. Chromosome painting in silico in a bacterial species reveals fine population structure. *Mol Biol Evol.* 2013;30:1454–64.
- Eppinger M, Baar C, Linz B, Raddatz G, Lanz C, Keller H, et al. Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *Plos Genet.* 2006;2:e120.
- Kersulyte D, Rossi M, Douglas B. Sequence divergence and conservation in genomes of *Helicobacter cetorum* strains from a dolphin and a whale. *Plos One.* 2013;8:e83177.
- Haesebrouck F, Pasmans F, Flahou B, Chiers K, Baele M, Meyn T, et al. Gastric helicobacters in domestic animals and non-human primates and their significance for human health. *Clin Microbiol Rev.* 2009;22:202–23.
- Dewhirst FE, Shen Z, Scimeca MS, Stokes LN, Boumenna T, Chen T, et al. Discordant 16S and 23S rRNA gene phylogenies for the genus *Helicobacter*: implications for phylogenetic inference and systematics. *J Bacteriol.* 2005;187:6106–18.
- Joosten M, Linden S, Rossi M, Tay ACY, Skoog E, Medea P, et al. Divergence between the highly virulent *Helicobacter heilmannii* and its closest relative, the low-virulence *Helicobacter ailurogastricus* sp. nov. *Infect Immun.* 2015;84:293–306.
- Flahou B, Haesebrouck F, Smet A. (2016). Non-*Helicobacter pylori* *Helicobacter* infections in humans and animals. *Helicobacter pylori Res.* 233–73.
- Gueneau P, De Goër SL. *Helicobacter*: molecular phylogeny and the origin of gastric colonization in the genus. *Infect Gen Evol.* 2002;1:215–33.
- Arnold AC, Zigova Z, Holden M, Lawley TD, Rad R, Dougan G, et al. Comparative whole genome sequence analysis of the carcinogenic bacterial model pathogen *Helicobacter felis*. *Genome Biol Evol.* 2011;3:302–8.
- O’Toole PW, Snelling WJ, Canchaya C, Forde BM, Hardie KR, Josenhans C, et al. Comparative genomics and proteomics of *Helicobacter mustelae*, an ulcerogenic and carcinogenic gastric pathogen. *BMC Genomics.* 2010;11:164.
- Schott T, Kondadi PK, Hänninen ML, Rossi M. Comparative genomics of *Helicobacter pylori* and the human-derived *Helicobacter bizzozeronii* CIII-1 strain reveal the molecular basis of the zoonotic nature of non-*pylori* gastric *Helicobacter* infections in humans. *BMC Genomics.* 2011;12:534.
- Smet A, Van Nieuwerburgh F, Ledesma J, Flahou B, Deforce D, Ducatelle R, et al. Genome sequence of *Helicobacter heilmannii* sensu stricto ASB1 isolated from the gastric mucosa of a kitten with severe gastritis. *Genome Announce.* 2013;1:e00033–12.
- Vermoote M, Vandekerckhove TT, Flahou B, Pasmans F, Smet A, De Groot D, et al. Genome sequence of *Helicobacter suis* supports its role in gastric pathology. *Vet Res.* 2011;42:51.
- Fox JG, Correa P, Taylor NS, Lee A, Otto G, Murphy JC, et al. *Helicobacter mustelae*-associated gastritis in ferrets: an animal model of *Helicobacter pylori* gastritis in humans. *Gastroenterology.* 1990;99:352–61.
- Fox JG, Marin RP. *Helicobacter mustelae* infection in ferrets: pathogenesis, epizootiology, diagnosis and treatment. *Sem Avian Exot Pet Med.* 2001;10:36–44.

26. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
27. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31:3691–3.
28. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol*. 2016;17:238.
29. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Met*. 2012;9:173–5.
30. Zimmerman L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M et al. (2017). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*. <https://doi.org/10.1016/j.jmb.2017.12.007>.
31. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
32. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucl Acids Res*. 2016;44:W242–245.
33. Cheng L, Connor TR, Siren J, Aanensen DM, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol Evol*. 2013;30:1224–8.
34. Sandeep JJ, Cox D, Wilff B, Morisson SS, Kozal-Muiznieks NA, Frace M, et al. Dynamics of genome change among *Legionella* species. *Sci Rep*. 2016;6:33442.
35. Corander J, Marttinen P. Bayesian identification of admixture events using multilocus molecular markers. *Mol Ecol*. 2006;15:2833–43.
36. Corander J, Marttinen P, Siren J, Tang J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinform*. 2008;9:539.
37. To TH, Jung M, Lycette S, Gascuel O. Fast dating using least-squares criteria and algorithms. *Syst Biol*. 2015;00:1–16.
38. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016;2:vew007.
39. Darch ES, McNally A, Harrison F, Corander J, Barr HL, Paszkiewicz, et al. Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. *Sci Rep*. 2015;5:7649.
40. Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, et al. *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci USA*. 2011;108:5033–8.
41. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. *Science*. 2007;315:476–80.
42. Walter J. Ecological role of Lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol*. 2008;74:4985–96.
43. Klemm E, Dougan R. Advances in understanding bacterial pathogenesis gained from whole genome sequencing and phylogenetics. *Cell Host Microbe*. 2016;19:599–610.
44. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proc Natl Acad Sci USA*. 2013;29:11923–7.
45. Lees JA, Vehkala M, Valimaki N, Harris SR, Chewapreecha C, Croucher NJ, et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat Commun*. 2016;7:12797.
46. Vinella D, Fischer F, Vorontsov E, Gallaud J, Malosse C, Michel V, Cavazza C, Robbe-Saule M, Richaud P, Chamot-Rooke J, Brochier-Armanet C, De Reuse H. Evolution of *Helicobacter*: acquisition by gastric species of two histidine-rich proteins essential for colonization. *Plos Path*. 2015;11:e10053.
47. Fischer F, Robbe-Saule M, Turlin E, Mancuso F, Michel V, Richaud P, et al. Characterization in *Helicobacter pylori* of a nickel transporter essential for colonization that was acquired during evolution by gastric *Helicobacter* species. *Plos Path*. 2016;12:e1006018.
48. Weinberg MV, Maier RJ. Peptide transport in *Helicobacter pylori*: roles of Dpp and Opp systems and evidence for additional peptide transporters. *J Bacteriol*. 2007;189:3392–402.
49. Oleastro M, Ménard A. The role of *Helicobacter pylori* outer membrane proteins in adherence and pathogenesis. *Biology*. 2013;2:1110–34.
50. Cynober L. Can arginine and ornithine support gut functions? *Gut*. 1994;35:S42–S45.
51. Baele M, Decostere A, Vandamme P, Ceelen L, Hellemans A, Mast J, et al. Isolation and characterization of *Helicobacter suis* sp. nov. from pig stomachs. *Int J Syst Evol Microbiol*. 2008;58:1350–8.
52. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, et al. Analysis of one million base pairs of Neanderthal DNA. *Nature*. 2006;444:330–6.
53. Nikaido M, Matsuno F, Hamilton H, Brownell RL, Cao Y, Ding W, et al. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc Natl Acad Sci USA*. 2001;98:7384–9.
54. Smith EM, Needs PF, Manley G, Green LE. Global distribution and diversity of ovine-associated *Staphylococcus aureus*. *Infect Genet Evol*. 2014;22:208–15.
55. Flahou B, Rossi M, Bakker J, Langermans J, Heuvelman E, Solnick J, et al. Evidence for a primate origin of zoonotic *Helicobacter suis* colonizing domesticated pigs. *ISME*. 2017. <https://doi.org/10.1038/ismej.2017.145>
56. O'Brien SJ, Johnson W, Driscoll C, Pontius J, Pecon-Slattery J, Menotti-Raymond M. State of cat genomics. *Trends Genet*. 2008;24:268–79.
57. Wayne RK, Ostrander EA. Lessons learned from the dog genome. *Trends Genet*. 2007;23:557–67.
58. De Witte C, Devriendt B, Flahou B, Bosschem I, Ducatelle R, Smet A, et al. *Helicobacter suis* induces changes in gastric inflammation and acid secretion markers in pigs of different ages. *Vet Res*. 2017;48:34.
59. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics*. 2017. <https://doi.org/10.1093/bioinformatics/btx610>